

Uncertainties in Interpolated Spectral Data

Volume 108

Number 1

January-February 2003

James L. Gardner

CSIRO National Measurement
Laboratory,
Lindfield, Australia 2070

jim.gardner@csiro.au

Interpolation is often used to improve the accuracy of integrals over spectral data convolved with various response functions or power distributions. Formulae are developed for propagation of uncertainties through the interpolation process, specifically for Lagrangian interpolation increasing a regular data set by factors of 5 and 2, and for cubic-spline interpolation. The interpolated data are correlated; these correlations must be considered when combining the interpolated values, as in integration. Examples are given using a common spectral integral in photometry.

Correlation coefficients are developed for Lagrangian interpolation where the input data are uncorrelated. It is demonstrated that in practical cases, uncertainties for the integral formed using interpolated data can be reliably estimated using the original data.

Keywords: interpolation; photometry; radiometry; uncertainty.

Accepted: December, 19, 2002

Available online: <http://www.nist.gov/jres>

1. Introduction

Measurements of spectral irradiance, spectral responsivity or spectral reflectance are often made for a limited set of wavelengths and then used to calculate weighted spectral sums for photometry, colorimetry or filter radiometry. It is often necessary to interpolate the spectral data to a finer grid to avoid errors arising from the discrete approximation used to estimate the integral where the weighting function varies strongly between the wavelengths at which the measurements were made [1]. Interpolated values are then correlated to the original data and to nearby interpolated points; unless these correlations are taken into account, uncertainty calculations will give misleading results, generally underestimating errors in the spectral sums by significant amounts.

Interpolation may also be required when reference standards are provided for a limited set of wavelengths. Many of the primary reference standards in the national metrology institutes are derived in a functional form, and can be calculated on as fine a wavelength grid as required, along with all the necessary correlations [2]. However, calibrations of client lamps or detectors involve a transfer, by comparison, to artifacts that do

not necessarily show a spectral variation that is easily modeled, and in the interests of reducing costs may be provided on a limited number of wavelengths. The reference data at different wavelengths may also be correlated. While the primary reference standards are often strongly correlated between wavelengths, the transfer process itself adds uncertainty that is generally random and often reduces the correlations to negligible levels [2]

Where data are available at sufficient wavelengths to avoid errors due to the sum approximation to the integral, it is preferable to interpolate reference tables, such as the photometric response function V_λ and the colorimetric tristimulus response functions [3] to the wavelengths at which measurements are taken, because those tables contain no uncertainty. However, interpolation of measurement data is often applied. One reason is that software for a calculation may require data at a particular interval, another is that instrument measurement programs may provide limited sets of data.

Following a brief description of uncertainty propagation, this paper is divided into two sections. The first covers Lagrangian interpolation, the second cubic spline interpolation; these are the two most commonly

used interpolation methods. In each section, interpolation from a data set which is itself correlated is considered. Various simplifications for practical applications are made, and examples are presented. A conclusion is that in practical terms, uncertainty can be accurately derived from the original data set without a complex calculation of correlations.

2. Propagation of Uncertainty

Uncertainty propagation is described in detail in the ISO *Guide to the Expression of Uncertainty in Measurement* [4]. The uncertainty in a quantity y formed by combining measured quantities x_i through the relationship $y = f(x_1, x_2, \dots, x_N)$ is given by

$$u^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + \sum_{i=1}^N \sum_{j \neq i=1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j), \quad (1)$$

where $u(x_i)$ is the uncertainty in x_i and $u(x_i, x_j)$ is the covariance between x_i and x_j . For uncorrelated input quantities, the covariance between pairs of variables is zero and Eq. (1) reduces to the "sum of squares" commonly applied. The derivatives $\partial f / \partial x_i$ are sensitivity coefficients for the dependence of y on the various measured quantities. Given that $u^2(x_i) \equiv u(x_i, x_i)$, Eq. (1) can be expressed as

$$u^2(y) = \mathbf{f}_y^T \mathbf{U}_x \mathbf{f}_y, \quad (2)$$

where

$$\mathbf{f} = \left(\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \dots \frac{\partial f}{\partial x_n} \right)^T \quad (3)$$

is a column vector of sensitivity coefficients (T indicates the transpose) and

$$\mathbf{U}_x = [u(x_i, x_j)] \quad (4)$$

is the $N \times N$ uncertainty matrix.

Interpolation of spectral data is generally performed to produce a set of values (p_k, x_k) from set (y_i, x_i) where x_i is the independent variable (generally wavelength in radiometry). Quantities in the set (p_k, x_k) that depend on the same (y_i, x_i) are correlated through this common dependence. The covariance between two values p_k and p_m (and, when $k = m$, the square of the uncertainty) is given by

$$u(p_k, p_m) = \sum_{i=1}^N \sum_{j=1}^N \frac{\partial p_k}{\partial y_i} \frac{\partial p_m}{\partial y_j} u(y_i, y_j). \quad (5)$$

In matrix form, this is simply expressed as

$$u(p_k, p_m) = \mathbf{f}_{p_k}^T \mathbf{U}_y \mathbf{f}_{p_m}. \quad (6)$$

It is sometimes convenient to use correlation coefficients rather than covariances, defined as

$$r(p_k, p_m) = \frac{u(p_k, p_m)}{\sqrt{u(p_k, p_k) u(p_m, p_m)}}. \quad (7)$$

A matrix of correlation coefficients is square, symmetric about the diagonal and has value 1 in the diagonal elements.

3. Lagrangian Interpolation

We have a tabulated function y_i at values x_i :

$$y_i = y(x_i), i = 1 \dots N. \quad (8)$$

For x in the range x_m to x_{m+n} , the formula for Lagrangian interpolation is [5]

$$\begin{aligned} p(x) &= \sum_{j=1}^n \frac{\prod_{i \neq j=1}^n (x - x_{m+i-1})}{\prod_{i \neq j=1}^n (x_{m+j-1} - x_{m+i-1})} y_{m+j-1} \\ &= \sum_{j=1}^n w_j y_{m+j-1} \end{aligned} \quad (9)$$

Equation (9) represents an $(n-1)$ th order polynomial fitted through the n original values. Interpolated data are formed as a linear combination of nearby existing data. Sensitivity coefficients for the dependence of the interpolated data on the input data are simply the weights w_j in Eq. (9). Covariances between the output values, that is, the original input values plus the interpolated values, take several forms. Correlations present in the original values remain. The values newly formed by interpolation are correlated to both the input values forming them (and through them to the remaining input values if correlations are present in the original set) and to any new values formed from common input values. All of these covariances, including the uncertainty of the interpolated values, can be calculated through Eq. (6).

In many instances, the input data y_i are present at regular values of x_i and further, the output data are also required at regular intervals. Interpolation is usually then performed by running Eq. (9) through the data and forming a new value or values in the center of the range. The multipliers convolving the data are deter-

mined by the interval spacings only and can be calculated prior to the interpolation. This was clearly demonstrated by Savitsky and Golay [6] in the development of their algorithms for smoothing and differentiation of spectral data where polynomial expressions of various order are fitted to regularly-spaced data; in those cases, the coefficients are determined as fixed linear combinations of the input dependent data. The arguments presented here for Lagrangian interpolation can easily be extended to cover smoothing of data using the Savitsky-Golay routines. Two-point Lagrangian interpolation forming a new value in the center of the existing values has weights $(w_1, w_2) = (1/2, 1/2)$, equivalent to a linear interpolation.

Propagation of uncertainty through two common examples of Lagrangian interpolation used in photometry are now discussed. In both of these we consider the calculation of illuminance response to CIE illuminant D65, a tabulated distribution carrying no uncertainty, for a photometer whose spectral response is a close approximation to V_λ , measured at different spectral intervals and where the measured response values are uncorrelated. These distributions are shown in Fig. 1 for a wavelength interval of 5 nm. As the response function tapers to zero at each end, the luminance response is given as

$$R_V = \Delta\lambda \sum_{i=1}^N R_i E_{D65,i}, \quad (10)$$

where R_i is the photometer response and $E_{D65,i}$ is the illuminant value at the i th wavelength, respectively, and $\Delta\lambda$ is the 5 nm wavelength separation between the values.

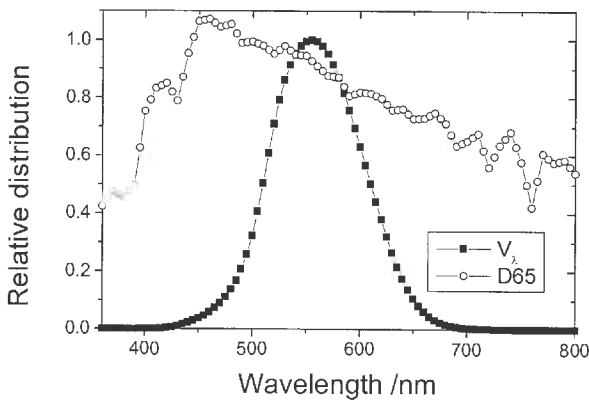


Fig. 1. Distributions CIE V_λ and D65 used for interpolation examples.

For uncorrelated spectral response values, the uncertainty in R_V is given by

$$u^2(R_V) = (\Delta\lambda)^2 \sum_{i=1}^N E_{D65,i}^2 u^2(R_i). \quad (11)$$

For values as tabulated by CIE [3], the value of R_V is 10567.41, with an uncertainty 19.60 (relative uncertainty 0.1855 %) if the responsivity values have a relative uncertainty of 1 % and are uncorrelated. (Note that extra significant figures for the uncertainty are presented above those of normal practice for the purpose of comparison.)

3.1 Photometer Measured at 10 nm Intervals, Interpolated to 5 nm

The spectral integral Eq. (10) for input values on a 10 nm grid evaluates to 10568.18, a small change compared to the 5 nm data due to the discrete approximation to the integral; the relative uncertainty for uncorrelated spectral response values with a relative uncertainty of 1 % becomes 0.2623 %, or the expected $\sqrt{2}$ increase compared to the response measured at 5 nm intervals. We wish to interpolate the photometer response with a four point Lagrangian function to data on a 5 nm interval. The weights for Eq. (9) are then

$$w_1, w_2, w_3, w_4 = \frac{-1}{16}, \frac{9}{16}, \frac{9}{16}, \frac{-1}{16} \quad (12)$$

and, as the input data are uncorrelated, the uncertainty for an interpolated value is given by

$$u^2(p) = \sum_{j=1}^4 w_j^2 u^2(y_{m+j-1}). \quad (13)$$

In any given interval spanning four input values, the uncertainty value $u(y_i)$ is approximately constant, and the interpolated values have an uncertainty approx. 80 % of the original values in that range. If we ignore the correlations that have been introduced, the relative uncertainty of the integral evaluates as 0.17 %, low compared to the original value and clearly incorrect as it would imply that it could be reduced to zero by repeated interpolation.

A full correlation matrix for the output values is formed as follows. First the interpolation is performed

(using linear interpolation in the first and last intervals of the input values). From N original values, we now have $(2N-1)$ values

$$\begin{aligned} y_i &= y_{\text{original},j}; i \text{ odd}, j = (i+1)/2 \\ y_i &= w_1 y_{i-3} + w_2 y_{i-1} + w_3 y_{i+1} + w_4 y_{i+3}; \\ 4 < i < 2N-4, i \text{ even}. \end{aligned} \quad (14)$$

Uncertainties for these values are known for the original data and for the interpolated data from Eq. (13) and these are used to populate the diagonal elements of the uncertainty matrix U_y . We then have to populate the elements to the right of the diagonal only before filling to the left of the diagonal by symmetry. As the original input values (now at i odd) are uncorrelated, we have for all i , from Eq. (5),

$$\begin{aligned} u(y_i, y_{i+1}) &= w_3 u^2(y_{i+1}) \\ u(y_i, y_{i+3}) &= w_4 u^2(y_{i+3}) \end{aligned} \quad (15)$$

For i even (interpolated values),

$$\begin{aligned} y_{i+2} &= w_1 y_{i-1} + w_2 y_{i+1} + w_3 y_{i+3} + w_4 y_{i+5} \\ y_{i+4} &= w_1 y_{i+1} + w_2 y_{i+3} + w_3 y_{i+5} + w_4 y_{i+7} \\ y_{i+6} &= w_1 y_{i+3} + w_2 y_{i+5} + w_3 y_{i+7} + w_4 y_{i+9} \end{aligned} \quad (16)$$

and from Eq. (5) we have

$$\begin{aligned} u(y_i, y_{i+2}) &= w_1 w_2 u^2(y_{i-1}) + w_2 w_3 u^2(y_{i+1}) + w_3 w_4 u^2(y_{i+3}) \\ u(y_i, y_{i+4}) &= w_1 w_3 u^2(y_{i+1}) + w_2 w_4 u^2(y_{i+3}) \\ u(y_i, y_{i+6}) &= w_1 w_4 u^2(y_{i+3}) \end{aligned} \quad (17)$$

For completeness, similar expressions were applied for the values formed by linear interpolation in the first and last intervals. Uncertainty calculated for the integral including all these correlations was then exactly that calculated with the original data for the 10 nm grid.

One further simplification can be made for uncorrelated input values. In practical terms, nearby values have the same uncertainty. Hence the sets of Eqs. (14) to (17) can be reduced to a matrix of correlation coefficients determined purely from the weights,

$$\begin{pmatrix} \dots & & & & & & \\ 1 & \frac{9}{\sqrt{164}} & \frac{63}{164} & \frac{-1}{\sqrt{164}} & \frac{-18}{164} & 0 & \frac{1}{164} \\ & 1 & \frac{9}{\sqrt{164}} & 0 & \frac{-1}{\sqrt{164}} & 0 & 0 \\ & & 1 & \frac{9}{\sqrt{164}} & \frac{63}{164} & \frac{-1}{\sqrt{164}} & \frac{-18}{164} \\ & & & 1 & \frac{9}{\sqrt{164}} & 0 & \frac{-1}{\sqrt{164}} \\ & & & & \dots & & \end{pmatrix}, \quad (18)$$

where the first row corresponds to an interpolated value. Uncertainty for the integral using the interpolated data set is then found by modifying the sensitivity column vector to include the uncertainty at each value,

$$f = \left(\frac{\partial f}{\partial y_i} u(y_i) \right), \quad (19)$$

and then performing the matrix multiplication Eq. (2). A negligible change relative to the true values is due to averaging through regions where the response is changing rapidly; while the relative uncertainty in these regions is constant, the absolute value is not.

Table 1 shows uncertainties calculated for all the options discussed in this section. It can be seen that proper accounting for the correlations introduced by the interpolation reproduces the uncertainty calculated for the input values alone, and that using the correlation coefficients provides a practical calculation of the uncertainty matrix.

Table 1. Integral value and relative uncertainty for various calculation options; four point Lagrange interpolation adding one value in each range. Input values uncorrelated, 1 % relative uncertainty.

Method	Integral	Relative uncertainty
Original data on 5 nm grid	10567.4	0.185
Original data on 10 nm grid	10568.2	0.262
Interpolated data; ignore correlations	10567.5	0.168
Interpolated data; with correlations	10567.5	0.262
Interpolated data; use correlation coefficients	10567.5	0.266

3.2 Photometer Measured at 5 nm Intervals, Interpolated to 1 nm

After re-arranging the N input values to their new positions, we add four values between input values y_i and y_{i+5} that can be represented as

$$\begin{pmatrix} y_{i+1} \\ y_{i+2} \\ y_{i+3} \\ y_{i+4} \end{pmatrix} = \frac{1}{125} \begin{pmatrix} -6 & 108 & 27 & -4 \\ -8 & 84 & 56 & -7 \\ -7 & 56 & 84 & -8 \\ -4 & 27 & 108 & -6 \end{pmatrix} \begin{pmatrix} y_{i-5} \\ y_i \\ y_{i+5} \\ y_{i+10} \end{pmatrix}. \quad (20)$$

Fig. 2 shows the relative uncertainty of the central data set where a photometer with a V_λ response is interpolated from 5 nm to 1 nm and the input values are assumed uncorrelated with a relative uncertainty of 1 %. Uncertainties of the interpolated values are lower than those of the input values. If we ignore the correlations introduced by the interpolation, the relative uncertainty of the integral with the D65 illuminant (interpolated to

a 1 nm grid with a cubic-spline routine), reduces to 0.074 %, a value too low by the order of $\sqrt{5}$. The integral itself has the same value as that shown in Table 1 for the 5 nm grid.

For a four point Lagrange interpolation adding four values between each of the input values, correlations in the output set extend over the 19 values following an input value. The matrix of correlation coefficients is shown as the transpose relative to Eq. (18) in the interests of printing; that is, it is equivalent to filling to the lower left of the diagonal of the correlation matrix prior to filling the upper right by symmetry. Beginning at a column corresponding to an input value, the matrix of correlation coefficients is

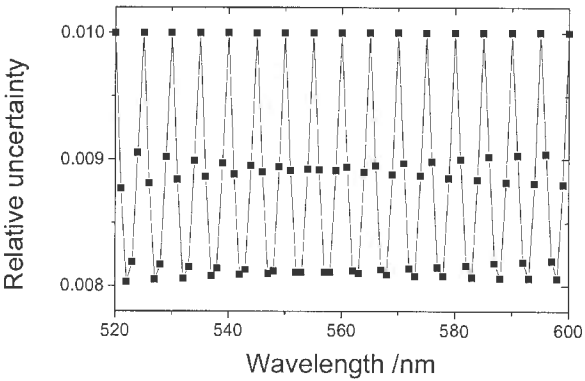


Fig. 2. Relative uncertainties for the interpolated V_λ data set. Original values assumed uncorrelated, with a relative uncertainty of 1 %.

1				
108	1			
$\sqrt{5 \cdot 2489}$				
84	2132			
$\sqrt{5 \cdot 2061}$	$\sqrt{2489 \cdot 2061}$	1		
56	1678	1678		
$\sqrt{5 \cdot 2061}$	$\sqrt{2489 \cdot 2061}$	$\sqrt{2061 \cdot 2489}$	1	
27	1176	1678	2132	
$\sqrt{5 \cdot 2489}$	2489	$\sqrt{2061 \cdot 2489}$	$\sqrt{2061 \cdot 2489}$	1
0	27	56	84	108
	$\sqrt{5 \cdot 2489}$	$\sqrt{5 \cdot 2061}$	$\sqrt{5 \cdot 2061}$	$\sqrt{5 \cdot 2489}$
27	318	861	1440	2016
$\sqrt{5 \cdot 2489}$	2489	$\sqrt{2061 \cdot 2489}$	$\sqrt{2061 \cdot 2489}$	2489
56	132	532	976	1440
$\sqrt{5 \cdot 2061}$	$\sqrt{2489 \cdot 2061}$	2061	2061	$\sqrt{2489 \cdot 2061}$
84	7	245	532	861
$\sqrt{5 \cdot 2061}$	$\sqrt{2489 \cdot 2061}$	2061	2061	$\sqrt{2489 \cdot 2061}$
108	-67	7	132	318
$\sqrt{5 \cdot 2489}$	2489	$\sqrt{2061 \cdot 2489}$	2489	2489
0	-4	-7	-8	-6
	$\sqrt{5 \cdot 2489}$	$\sqrt{5 \cdot 2061}$	$\sqrt{5 \cdot 2061}$	$\sqrt{5 \cdot 2489}$
0	-114	-210	-264	-252
	2489	$\sqrt{2061 \cdot 2489}$	$\sqrt{2061 \cdot 2489}$	2489
0	-104	-196	-256	-264
	$\sqrt{2489 \cdot 2061}$	2061	2061	$\sqrt{2489 \cdot 2061}$
0	-77	-147	-196	-210
	$\sqrt{2489 \cdot 2061}$	2061	2061	$\sqrt{2489 \cdot 2061}$
0	-40	-77	-104	-114
	2489	$\sqrt{2061 \cdot 2489}$	$\sqrt{2061 \cdot 2489}$	2489
0	0	0	0	0
0	24	42	48	36
	$5 \cdot 2489$	$5 \cdot \sqrt{2061 \cdot 2489}$	$5 \cdot \sqrt{2061 \cdot 2489}$	$5 \cdot 2489$
0	32	56	64	48
	$5 \cdot \sqrt{2489 \cdot 2061}$	$5 \cdot 2061$	$5 \cdot 2061$	$5 \cdot \sqrt{2489 \cdot 2061}$
0	28	49	56	42
	$5 \cdot \sqrt{2489 \cdot 2061}$	$5 \cdot 2061$	$5 \cdot 2061$	$5 \cdot \sqrt{2489 \cdot 2061}$
0	16	28	32	24
	$5 \cdot 2489$	$5 \cdot \sqrt{2061 \cdot 2489}$	$5 \cdot \sqrt{2061 \cdot 2489}$	$5 \cdot 2489$
0	0	0	0	0

(21)

A number of these, for the values furthest from the input set, are negligible. The relative uncertainty of the integral for the interpolated data set calculated with these correlation coefficients is 0.197 %, equivalent in practical terms to that calculated using the original 5 nm data set and shown in Table 1.

4. Cubic-Spline Uncertainty Propagation

For our set of data Eq. (8), cubic-spline interpolation [7] calculates a value y at x in the interval x_i to x_{i+1} as

$$y = Ay_i + By_{i+1} + Cy_i'' + Dy_{i+1}'', \quad (22)$$

where

$$A = \frac{x_{i+1} - x}{x_{i+1} - x_i}, \quad (23)$$

$$B = \frac{x - x_i}{x_{i+1} - x_i}, \quad (24)$$

$$C = \frac{1}{6}(A^3 - A)(x_{i+1} - x_i)^2, \quad (25)$$

$$D = \frac{1}{6}(B^3 - B)(x_{i+1} - x_i)^2. \quad (26)$$

The first two terms of Eq. (22) represent simple linear interpolation. Including the second derivatives y'' yields a function that has first and second derivatives continuous at the boundaries between intervals.

The second-derivatives are unknown. The relation between them is given by

$$\begin{aligned} \frac{x_i - x_{i-1}}{6} y_{i-1}'' + \frac{x_{i+1} - x_{i-1}}{3} y_i'' + \frac{x_{i+1} - x_i}{6} y_{i+1}'' = \\ \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - \frac{y_i - y_{i-1}}{x_i - x_{i-1}}, \end{aligned} \quad (27)$$

which is a system of $N-2$ equations in the N unknowns y_i'' . The natural cubic-spline, which is commonly used, sets

$$y_i'' = y_N'' = 0 \quad (28)$$

and solves for the remaining terms [7]. We are interested in using the cubic-spline interpolation on spectral data of known uncertainties, including the possibility of correlations, where the interpolated data may then be combined in various ways, so that not only the uncertainties in the interpolated data but also the correlations present are important in propagating uncertainties in the combinations.

The $(N-2)$ values of y_i'' depend on each of the input values, i.e., are correlated to each input value.

Then even for uncorrelated input data, the output data are correlated over the whole set of interpolated values. We wish to calculate the covariance $u(y_n, y_m)$ between two interpolated values

$$\begin{aligned} y_m &= A_m y_i + B_m y_{i+1} + C_m y_i'' + D_m y_{i+1}'' \\ y_n &= A_n y_j + B_n y_{j+1} + C_n y_j'' + D_n y_{j+1}'' \end{aligned}, \quad (29)$$

where the y_n and y_m values may be in the same or different intervals denoted by i, j . The uncertainty in y_n is given by

$$u^2(y_n) = u(y_n, y_n). \quad (30)$$

The covariance between (and uncertainty of) the input values is known, carried in the matrix

$$U_y = [u(y_i, y_j)]. \quad (31)$$

To propagate uncertainties through Eq. (29) we need the covariance between the second-derivative values and the input values, and between the second-derivatives themselves. These in turn require the sensitivity coefficients for the dependence of the second-derivative on the input values, $\partial y_k'' / \partial y_i$.

The set of $N-2$ equations Eq. (27) can be written in matrix form

$$\begin{pmatrix} \frac{y_3 - y_2}{x_3 - x_2} - \frac{y_2 - y_1}{x_2 - x_1} & & & \\ & \ddots & & \\ \frac{y_{i+2} - y_{i+1}}{x_{i+2} - x_{i+1}} - \frac{y_{i+1} - y_i}{x_{i+1} - x_i} & & & \\ & \ddots & & \\ \frac{y_N - y_{N-1}}{x_N - x_{N-1}} - \frac{y_{N-1} - y_{N-2}}{x_{N-1} - x_{N-2}} & & & \end{pmatrix} = (h_{i,j}) \begin{pmatrix} y_2'' \\ \vdots \\ y_{N-1}'' \end{pmatrix}, \quad (32)$$

where

$$\begin{aligned} h_{i,i-1} &= \frac{x_i - x_{i-1}}{6} \\ h_{i,i} &= \frac{x_{i+2} - x_i}{3} \\ h_{i,i+1} &= \frac{x_{i+2} - x_{i+1}}{6} \end{aligned} \quad (33)$$

is a tri-diagonal matrix with remaining terms zero. Multiplying both sides of Eq. (32) on the left by the inverse matrix $(h_{i,j})^{-1}$ yields the required second-

derivatives. By selecting the i th row of this multiplication we have the relationship

$$y_{i+1}'' = \sum_{j=1}^{N-2} h_{i,j}^{-1} \left(\frac{y_{j+2} - y_{j+1}}{x_{j+2} - x_{j+1}} - \frac{y_{j+1} - y_j}{x_{j+1} - x_j} \right), \quad (34)$$

which when rearranged to

$$\begin{aligned} y_{i+1}'' &= \sum_{j=3}^N h_{i,j-2}^{-1} \frac{y_{j+2}}{x_{j+2} - x_{j+1}} \\ &\quad - \sum_{j=2}^{N-1} h_{i,j-1}^{-1} \left(\frac{1}{x_{j+2} - x_{j+1}} \right) + \left(\frac{1}{x_{j+1} - x_j} \right) y_j \\ &\quad + \sum_{j=1}^{N-2} h_{i,j}^{-1} \frac{y_j}{x_{j+1} - x_j} \end{aligned} \quad (35)$$

yields

$$\begin{aligned} \frac{\partial y_{i+1}''}{\partial y_j} &= \frac{h_{i,j-2}^{-1}}{x_j - x_{j-1}} \\ &\quad - h_{i,j-1}^{-1} \left(\frac{1}{x_{j+1} - x_j} + \frac{1}{x_j - x_{j-1}} \right) \\ &\quad + \frac{h_{i,j}^{-1}}{x_{j+1} - x_j}. \end{aligned} \quad (36)$$

These are augmented by

$$\partial y_k'' / \partial y_i = 0, k = 1, N. \quad (37)$$

If the input y_i values are uncorrelated, the required covariances are given by

$$u(y_i'', y_j'') = \sum_{k=1}^N \frac{\partial y_i''}{\partial y_k} \frac{\partial y_j''}{\partial y_k} u^2(y_k) \quad (38)$$

and

where

$$U = \begin{pmatrix} u^2(y_i) & u(y_{i+1}, y_i) & u(y_i'', y_i) & u(y_{i+1}'', y_i) & u(y_j, y_i) & u(y_{j+1}, y_i) & u(y_j'', y_i) & u(y_{j+1}'', y_i) \\ & u^2(y_{i+1}) & u(y_i'', y_{i+1}) & u(y_{i+1}'', y_{i+1}) & u(y_j, y_{i+1}) & u(y_{j+1}, y_{i+1}) & u(y_j'', y_{i+1}) & u(y_{j+1}'', y_{i+1}) \\ & & u^2(y_i'') & u(y_{i+1}'', y_i'') & u(y_j, y_i'') & u(y_{j+1}, y_i'') & u(y_j'', y_i'') & u(y_{j+1}'', y_i'') \\ & & & u^2(y_{i+1}'') & u(y_j, y_{i+1}'') & u(y_{j+1}, y_{i+1}'') & u(y_j'', y_{i+1}'') & u(y_{j+1}'', y_{i+1}'') \\ & & & & u^2(y_j) & u(y_{j+1}, y_j) & u(y_j'', y_j) & u(y_{j+1}'', y_j) \\ & & & & & u^2(y_{j+1}) & u(y_j, y_{j+1}) & u(y_{j+1}'', y_{j+1}) \\ & & & & & & u^2(y_j'') & u(y_{j+1}'', y_j'') \\ & & & & & & & u^2(y_{j+1}'') \end{pmatrix} \quad (45)$$

$$\begin{aligned} u(y_i'', y_j'') &= \sum_{k=1}^N \frac{\partial y_i''}{\partial y_k} \frac{\partial y_j''}{\partial y_k} u^2(y_k) \\ &= \frac{\partial y_i''}{\partial y_k} u^2(y_k) \end{aligned} \quad (39)$$

Note that the $N \times N$ matrix represented by Eq. (39) is not symmetric. For correlated inputs, the sums required are the matrix products

$$\begin{aligned} u(y_i'', y_j'') &= \mathbf{f}_i^T \mathbf{U}_y \mathbf{f}_j \\ u(y_i'', y_j'') &= \mathbf{f}_i^T \mathbf{U}_y \mathbf{g}_j \end{aligned}, \quad (40)$$

where

$$\mathbf{f}_i = \left(\frac{\partial y_i''}{\partial y_k} \right), k = 1..N \quad (41)$$

is a column-vector of sensitivity coefficients of the second-derivatives vs the input values and \mathbf{g}_j is a column vector of length N with 1 in the j th row, 0 elsewhere.

Covariances and uncertainties for the interpolated values are then found by recognizing that the values of y_m and y_n in Eq. (29) are represented by

$$\begin{aligned} y_m &= \mathbf{f}_m (y_i \ y_{i+1} \ y_i'' \ y_{i+1}'' \ y_j \ y_{j+1} \ y_j'' \ y_{j+1}'')^T \\ y_n &= \mathbf{f}_n (y_i \ y_{i+1} \ y_i'' \ y_{i+1}'' \ y_j \ y_{j+1} \ y_j'' \ y_{j+1}'')^T \end{aligned} \quad (42)$$

where

$$\begin{aligned} \mathbf{f}_m &= (A_i \ B_i \ C_i \ D_i \ 0 \ 0 \ 0 \ 0) \\ \mathbf{f}_n &= (0 \ 0 \ 0 \ 0 \ A_j \ B_j \ C_j \ D_j) \end{aligned} \quad (43)$$

are sensitivity vectors for y_m and y_n against the eight variables in the vectors shown in Eq. (42). Covariances between the interpolated values are then given by

$$u(y_m, y_n) = \mathbf{f}_m^T \mathbf{U} \mathbf{f}_n, \quad (44)$$

with the lower half symmetric about the diagonal. Only half of this matrix is required. The two right quadrants can be separately multiplied on the right by the column vector $(A_j \ B_j \ C_j \ D_j)^T$ and the two results combined into a single eight-element column vector for the final multiplication.

4.1 Cubic-Spline Interpolation Examples

Consider again the photometer response curve of Fig.1, to be integrated over the wavelength range from 360 nm to 830 nm (effectively the photometric response to an equal-energy source). The function was interpolated over the same input range (one less value) but shifted by 2.5 nm and the integral recalculated. Similarly, the function was interpolated to 1 nm intervals and the integral recalculated. Table 2 shows the results for these interpolations, where the uncertainty in the integral was calculated based on 1 % uncertainty in the input values, uncorrelated between values. The consequence of ignoring the correlation between the interpolated values is also shown. Correlations between distant points, introduced through the dependence of the second-derivatives, were negligible (largely because the response curve is relatively smooth), but strong correlations were found between near-neighbours.

Table 2. Integral of V_λ and its uncertainty for input values on a 5 nm grid and 1 % relative uncertainty, uncorrelated

	Integral	Uncertainty	Uncertainty ignoring correlations
Original data	106.8561	0.19647	
Shift 2.5 nm	106.8559	0.19647	0.1459
Interpolate to 1 nm	106.8559	0.19647	0.0746

Figure 3 shows propagated uncertainties for the interpolation shifting the input by 2.5 nm; for an interpolation to the mid-point, we would expect the interpolated value for a smooth function to be near the mean of the two interval boundaries with a propagated uncertainty $1/\sqrt{2}=0.71$ of the input (but of course correlated to adjacent values).

Figure 4 shows the variation in uncertainty for values interpolated at different positions within the 5 nm interval. This is a practical concern where a wavelength offset may be present in measurements, although in general it is a better practice to retain the wavelength values of the measured points and interpolate weighting functions such as the illuminant or the colorimetric response

functions as these carry no uncertainty. At the input points, uncertainties equal that of the input; at the mid-points they fall to $1/\sqrt{2}$ of the input points. Again this is expected as the input data are smooth, and cubic-spline results are not much different from linear interpolation. The cubic-spline reproduces the input ordinate values for abscissa values equal to an input value; for these points, we expect the propagated uncertainty to be that of the input and correlations between similar such values to match that of the input matrix U_y . These conditions can be used to test the coding.

Figure 5 shows V_λ interpolated from a 20 nm grid to 2 nm. Where the input curve is changing rapidly relative to the magnitude of the data, linear interpolation would be discontinuous and for these regions, the relative uncertainty of the interpolated values rises above that of the 1 % assumed for the input values. This is shown more strongly in Fig. 6 for input data spaced at 40 nm, interpolated to 5 nm. Here the interpolation does not provide a good representation of V_λ (as shown in Fig. 7); where the interpolation is poor, the uncertainties for the interpolated values rise above those of the input.

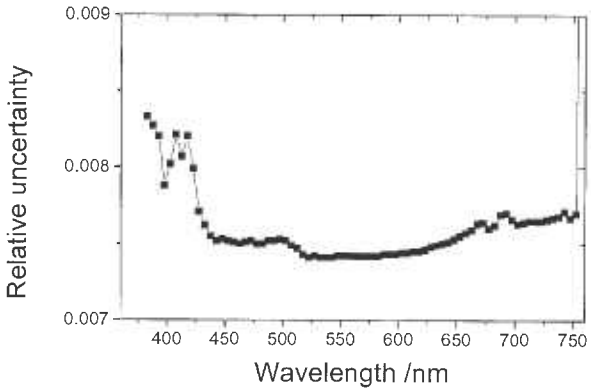


Fig. 3. Propagated uncertainty of V_λ for values shifted 2.5 nm; relative uncertainty of the input values is 1 %.

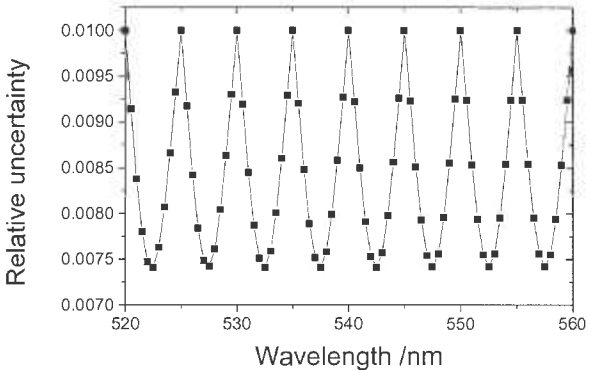


Fig. 4. Relative uncertainty of interpolated values for V_λ vs position in the interval. Input values with a relative uncertainty of 1 % are separated by 5 nm.

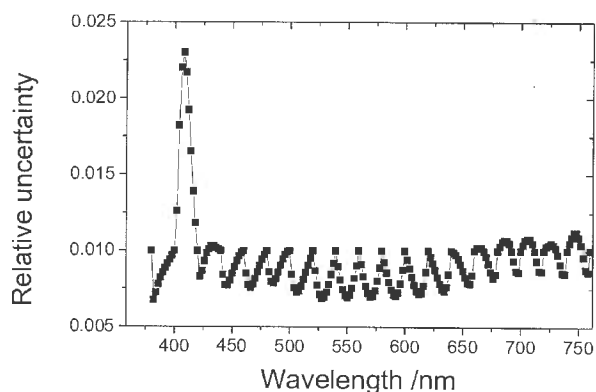


Fig. 5. Relative uncertainty for V_λ interpolated from a 20 nm grid to 2 nm. Relative uncertainty of the input values set to 1 %.

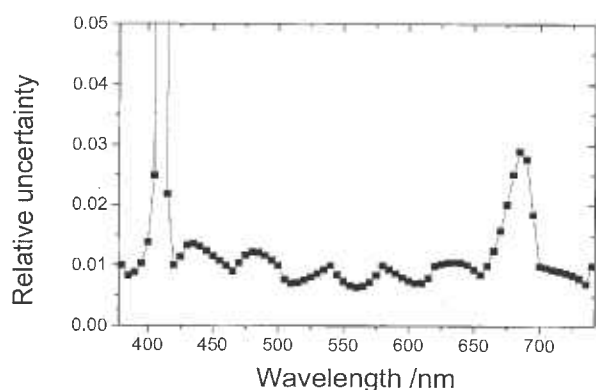


Fig. 6. Relative uncertainty for V_λ interpolated from a 40 nm grid to 5 nm. Relative uncertainty of the input values set to 1 %.

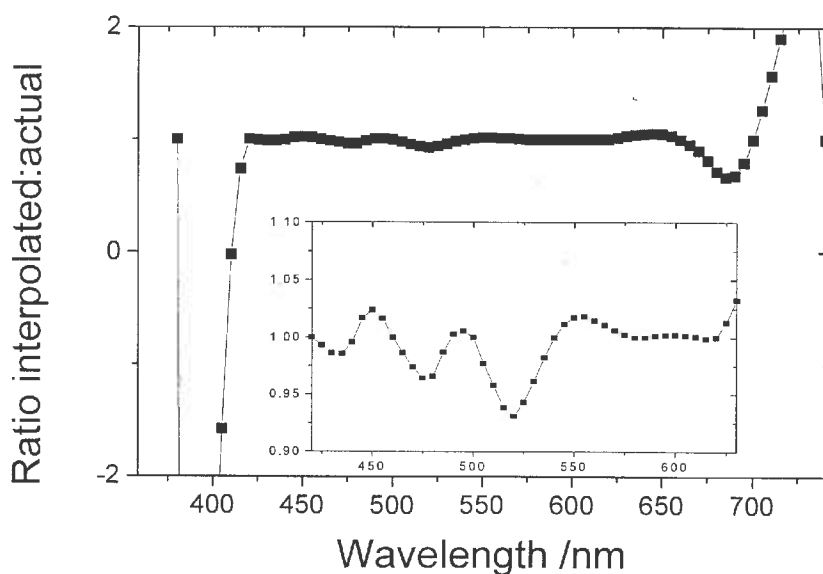


Fig. 7. Ratio of V_λ interpolated from a 40 nm grid to the actual value on a 5 nm grid. The inset shows the central region on an expanded scale.

5. Conclusion

Interpolation of spectral data is a common occurrence in radiometric and photometric measurements. Those data are often then combined in forming integral values such as photometric or colorimetric responses or filter radiometer responses. Interpolation is particularly important when a relatively smooth curve available only on a wide spectral spacing may need to be convolved with a more-rapidly changing curve and then integrated. Uncertainties in the interpolated values will generally be smaller than those of the input data, although this is not always true in the case of cubic-

spline interpolation. Uncertainties in combinations calculated using the interpolated data set then must include correlations introduced by the interpolation. Ignoring the correlation will lead to significant underestimation of uncertainties. The calculations in this paper for both Lagrange interpolation and for cubic-spline interpolation show that the uncertainty can be reliably estimated, in practical terms, by propagating the uncertainty through the combination using only the original set of data.

6. References

- [1] ASTM E 308-9, Standard practice for computing the colors of objects by using the CIE system, American Society for Testing and Materials, PA (2000).
- [2] J. L. Gardner, Correlations in primary spectral standards, Metrologia, to be published (NEWRAD 2001 proceedings).
- [3] CIE 15.2, Colorimetry, International Commission on Illumination, Vienna (1986).
- [4] Guide to the expression of uncertainty of uncertainty in measurement, International Organization for Standardization, Geneva (1993).
- [5] Numerical Recipes in C: the art of scientific computing, Cambridge Press (1992) p. 108.
- [6] A. Savitsky and M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. **36**, 1627-1639 (1964).
- [7] Numerical Recipes in C: the art of scientific computing, Cambridge Press (1992) p. 113.

About the author: James L. Gardner is a research fellow with the CSIRO National Measurement Laboratory, Sydney, Australia and was a Guest Researcher at NIST in the Optical Technology Division of the NIST Physics Laboratory. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.